Merkel, Uwe

Distributed TELEMAC parallel computing for small and medium enterprise infrastructure using the “Telemac-LinuxLiveCluster”

Distributed TELEMAC parallel computing for small and medium enterprise infrastructure using the "Telemac-LinuxLiveCluster"

Uwe H. Merkel *(Author)*
Ingenieurbüro Merkel, UHM
Karlsruhe, Germany
info@uwe-merkel.com

***Abstract* — A It is difficult and time consuming for both, beginners and advanced users, to set up a Telemac environment with parallel computing (distributed MPI-cluster) functionality and the many powerful Open Source pre- and postprocessing tools like PARAVIEW or QGIS. This article describes a ready to use Linux Life Disk that builds up a basic parallel computing cluster within few minutes. It lowers the hurdles of linux configuration and lets the user focus on the hydraulic analysis instead.**

Multiple copies of the disc can be plugged in existing office computers via USB or SATA and interconnect through common LAN infrastructure which is already available in most small and medium enterprises.

I. INTRODUCTION

Setting up a running system with Telemac, pre- and post-processing tools and all libraries is very time consuming. Setting up a Linux cluster for parallel computing is even worse. Before a user can care about his hydraulic projects, he needs weeks of trial and error to get all the right versions together. This is a big barrier for beginners and an expensive adventure for occasional users.

Therefore the Telemac Cluster abilities are only used by bigger institutions, as they have the human resources to care for the administration. This barrier was also the key problem for Telemac trainings, as many users had problems to repeat the exercises on their home computer setup.

The here presented project breaks down this barrier:

One of the biggest strengths of Linux is the ability of a 1:1 duplicated hard disk (with an installed and configured Linux) to be plugged in almost any other recent computer, and it runs out of the box, without reconfiguration. All the previous installed Telemac and cluster related special software runs as well.

The "Telemac-LinuxLiveCluster" makes use of the well known LinuxMint / Ubuntu LTS distributions which have an immense internet knowledge base and a reliable update cycle. The ready configured disk image, which contains free software for the full tool chain from pre- to post-processing, is free. It aims to provide a full working environment, running out of the box, for beginners as well as for hydraulic modelers who want to use development environments and cluster capabilities in a small office.

Plugging a 1:1 copy of the disk in any recent laptop or workstation integrates it in the local office cluster after just changing the IP address.

II. FEATURES

- TELEMAC Suite (several versions), precompiled for productive, debugging and profiling use, tested in parallel.
- PARAVIEW with a direct Selafin reader for Telemac binary files.
- QGIS with a direct Selafin reader for Telemac binary files and the TRIANGLE based mesh generator BASE MESH.
- A special GDAL beta version that converts Selafin files to docents of other vector formats like ESRI SHAPE, SVG, DXF, SQLITE, GeoJSON, GPX, KML a.o.
- Microsoft EXCEL and a Libre Office CALC files with macros able to read directly values from Selafin files into tables.
- TorquePBS, a scheduler that distributes large numbers of Telemac jobs to free cluster nodes, if necessary only at night or weekends.
- Cluster administration tools.
- Code::Blocks IDE, Fortran Edition with preconfigured Telemac setup to develop new princi.f files with several compilers (GNU - installed, NAG and INTEL - preconfigured).
- BACKInTIME, an automated backup system similar to Apples "Time Machine".

- WINE, a windows emulation environment, preconfigured to install BLUE KENUE (a Windows 64bit program) with a few clicks under LINUX.

Additionally the user might combine several instances of the workstation to a cluster by just duplicating the disk for other computers in the local area network.

Plugging the disks in via USB3.0 or SATA provides good performance. A basic 1-GBit switch is enough for at least 8 computers.

Tests with 40 cores on 8 Desktop computers showed similar performance as big clusters running on the same number and type of CPU cores:

- A speed up of more then x20 for the MPI parallel computation against serial computation of the official Telemac validation case “221_donau”
- A combined batch processing speedup of x35 for a 1000 job Telemac2d parameter study, compared to single core

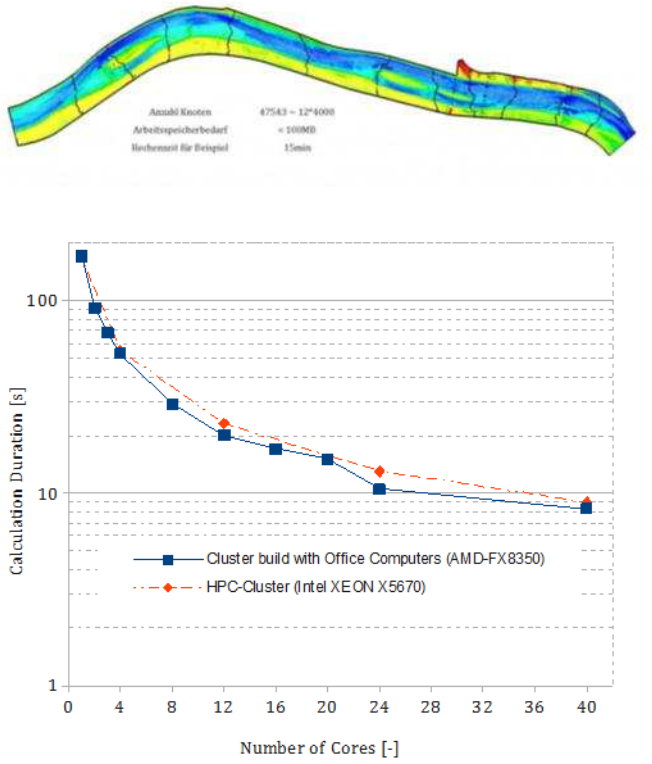

Figure 1. Parallel performance scaling ability of 150€ gaming CPUs with 1GB-LAN vs. 1300€ server CPUs.

III. ADVANTAGES & LIMITATIONS

The guideline for the development of this disc was: “It shall be a system for quick implementation and faster work flow on small and medium office networks with standard hardware, for demonstration and training purposes”. It was optimized and tested with Telemac projects divided in up to 40 parallel partitions on 8 computers with 40 cores. All nodes where average Desktop computers.

It is meant to run without much administration costs and complicated cluster safety policies.

The “Telemac-LinuxLiveCluster” scales up well as long as limit of the existing network infrastructure is not reached and the different nodes are not too heterogeneous in their performance.

The configuration is not optimized for network safety! (neither firewall nor virus protection are implemented for performance and cluster interaction reasons)

The IP-addresses are fixed, and DHCP is deactivated, as MPI will not run reliable with dynamic IP addresses. Currently the addresses 192.168.178.71 to 192.168.178.82 are used, but this can easily be changed in the file /etc/network/interfaces and the script /root/duplicate_node.sh

IV. LICENSE

The disc image of the “Telemac-LinuxLifeCluster” is available for free on demand by our website for everybody interested. It is free of charge and free of any warranty. It is meant as a quick start example for training purposes, for production use every user should set up his own system. Based on Ubuntu 12.04 (“Long Time Support” until 2017) and Linux Mint 13 all additionally installed programs are either under GNU and LPGL licenses that allow redistribution, or redistribution is directly granted by the authors.

V. SETUP

A. *The Server - Node71*

The disk is available as an image file at

www.uwe-merkel.com/TelemacLinuxLiveCluster and can be copied on any USB3.0-key with more than 60GB or much better, an internally connected solid state disc. The disc should be reasonable fast for a smooth workflow during pre- and postprocessing. One line in the Linux bash writes the image to a new disc, which will be completely overwritten. As superuser type:

```
gunzip -c clusterdisc.img.gz | dd
of=/dev/<the_new_disc>
```

Plug the new disc in any computer and restart it, and choose this disc as boot device in the BIOS (mostly with DEL or F12-buttons during start up)

During the first startup the system reconfigures itself for the new hardware what might take some time. When the system shows up, login with “guest” and enter the password “guest”.

Now a full functional Telemac Workstation is already running. The user is strongly advised to change the login and

password as soon as possible. The Linux Mint and Ubuntu online manuals provide help if needed.

B. The Clients

To set up a cluster together with other computers in the same network requires more copies of the same disc. This copies can be produced the same way as the first disc "Server" or by copying the disc directly with

```
dd if=/dev/sdb bs=16M| pv -s 60G | dd of=/dev/sdc
bs=16M
```

where /dev/sdb has to be replaced with the source disc and /dev/sdc has to be replaced with the next new disc.

Copying the disc is only possible from an external operating system. Use any other Linux computer or choose on boot up of the "Telemac-LinuxLiveCluster" the "Bootable ISO Image: pmagic". The latter option boots into a "parted magic rescue system", a tiny Linux that fits completely in a ram disc, from where it allows save copying of the "Telemac-LinuxLiveCluster".

Now each node needs an individual name and IP-address. This is done by a script which has to be modified and executed only once. After the first startup of each node open a terminal window (the BASH console) and login as superuser:

```
$> su                          #login as super user

$> cd /root/   #change to the root home

$> emacs duplicate_node.sh     #edit the config script

    # In the editor go to the following line and set
    # a unique node ID between 72 and 81 :

    nodenumber_new=$(echo '72')

    # Save the file

$> source duplicate_node.sh    #run the script

    #Output should look similar to this:

    This Client is node72:
    Hostname: node72
    Grub.cfg: 'node72 - Linux Mint 13 MATE'
    Fixed IP:  192.168.178.72

    This Clients network interface:
    EthernetDevice: auto eth1
    iface eth1 inet static


    The Server is node71:
    cluster-data : node71:/data  /cluster-data    nfs
    Torque Master: node71


$> reboot                      #reboots the computer
```

After rebooting, the client is a full Telemac workstation as well which can be used stand alone from the screen or remote controlled via SSH from any other client or server.

The SSH login is password free via the command:

```
ssh node72
```

ClusterSSH is able to login on various nodes at the same time, e.g.

```
cssh node71 node72 node73
```

All clients use the network folder */cluster-data* , which is placed on the server for jobs send to them by MPI or the Torque PBS-Queuing.

More Information can be found on the website.

C. Connecting the Clients (Hardware)

For small clusters the network hardware is only of minor importance. But it is strongly advised to use Ethernet cables at least in quality CAT6a or above, as they connect with up to 120MB/s to the other nodes. The switch is less important as long as you connect only up to 8 nodes. For more nodes a *managed switch* is mandatory.

D. Performance

Just connecting more computers will not necessarily bring a speed up for Telemac in parallel. The whole system will not be faster in calculating one time step than the slowest client! (as the neighboring clients need the result of the latter after each time step!) This means: All nodes should run on more or less the same performance level!

Despite of this, the cluster performance is mainly dominated by the following hardware parameters: (from "o = not very important" to "+++ = very important")

```
Server / Client:

  RAM:
    RAM-Frequency (+++)
    Number of RAM modules: double or
quadchannel(+++)
    Amount of RAM in GB (o/+)

  CPU:
    CPU-Frequency (+)
    Number of floating Point Units (+++)
    Number of Cores (++/+)
    Cache (+++)

  DISK:
    Writing performance (use SSDs !) (++)

  GPU: (o)
    not used for  Telemac, but significant in
PARAVIEW, known      problems with NVIDIA graphic

  Mainboard:
    Frequency (++)

Network:
    1GB-LAN is enough for up to 8 nodes with 4 cores
(o/+)
    cable: at least quality level "Cat6a" (++)
    1GB-Switch (+ / ++)
```

ACKNOWLEDGEMENT

Thanks to the help of many members of the Telemac Forum, members of the Telemac consortium, the feedback of costumers and participants of the trainings, the Telemac-LinuxLiveCluster could be brought to a level which is ready for every day usage.